

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently amended) A method for optimizing a database management system process of a query, the method comprising:
 - collecting a plurality of single column statistics for a plurality of columns, the plurality of single column statistics providing an estimate of row counts and unique entry counts for a single column operator;
 - selecting a preferred single column statistic from the plurality of single column statistics according to a predetermined criteria;
 - storing the preferred single column statistic;
 - determining a selectivity estimate for predicates in the query using the preferred single column statistic and a cross product of row counts for two columns selected from the plurality of columns, the selectivity estimate being used in optimizing processing of the query by the database management system.
2. (Original) The method of claim 1, wherein the predetermined criteria is a maximum of unique entry counts.
3. (Currently amended) The method of claim 2, further comprising:
 - ~~determining a cross product from the single column statistics;~~ and
 - calculating the selectivity estimate as the division of the cross product and the maximum of unique entry counts.
4. (Original) The method of claim 1, wherein the plurality of single column statistics are selectivities.

5. (Original) The method of claim 4, wherein the predetermined criteria is a minimum of selectivities.

6. (Currently amended) The method of claim 5, further comprising:
~~determining a cross product from the single column statistics; and~~
calculating the selectivity estimate as the product of the minimum of selectivities and the cross product.

7. (Original) The method of claim 1, wherein the plurality of columns are dependent on each other.

8. (Currently amended) A method for optimizing a database management system process of a query, the method comprising:

collecting a plurality of single column statistics for a plurality of columns,
the plurality of single column statistics providing an estimate of row counts and unique entry counts for a single column operator;

selecting a first preferred single column statistic from the plurality of single column statistics according to a first predetermined criteria;

determining a second preferred single column statistic from a first relationship of the single column statistics;

storing the first and second preferred single column statistics;

determining a selectivity estimate for predicates in the query using the first and second preferred single column statistics and a cross product of row counts for two columns selected from the plurality of columns, the selectivity estimate being used in optimizing processing of the query by the database management system.

9. (Original) The method of claim 8, wherein the first predetermined criteria is a maximum of unique entry counts.

10. (Currently amended) The method of claim 8, further comprising:
— ~~determining a cross product from the single column statistics; and~~
calculating the selectivity estimate as the division of the cross product and the maximum of unique entry counts.
11. (Original) The method of claim 8, wherein the first relationship of the single column statistics is a product of single column statistics.
12. (Original) The method of claim 8, wherein the plurality of single column statistics are selectivities.
13. (Currently amended) The method of claim 12, further comprising:
— ~~determining a cross product from the single column statistics; and~~
calculating the selectivity estimate as the product of the minimum of selectivities and the cross product.
14. (Original) The method of claim 12, wherein the first predetermined criteria is a minimum of selectivities.
15. (Original) The method of claim 8, wherein the plurality of columns are dependent on each other.
16. (Original) The method of claim 8, wherein the selectivity estimate is within a range between the first and second preferred single column statistics.
17. (Original) The method of claim 8, wherein the plurality of columns are substantially independent of each other.
18. (Original) The method of claim 17, wherein the selectivity estimate is substantially equal to the first preferred single column statistic.

19. (Original) The method of claim 8, wherein the columns are substantially dependent on each other.

20. (Original) The method of claim 19, wherein the selectivity estimate is substantially equal to the second preferred column statistic.

21. (Currently amended) A method for optimizing a database management system process of a query, the method comprising:

collecting a plurality of single column statistics for a plurality of columns,
the plurality of single column statistics providing estimates for row
counts and unique entry counts for a single column operator;

determining a first selectivity estimate based on an assumption that the
columns are substantially independent of each other;

determining a second selectivity estimate based on an assumption that
the columns are substantially dependent on each other;

determining a third selectivity estimate for predicates in the query using
the first and second selectivity estimates, the third selectivity
estimate being used in optimizing processing of the query by the
database management system.

22. (Original) The method of claim 21, further comprising:

determining a cross product from the single column statistics;

determining a measure of dependency; and

calculating the selectivity estimate as the product of a difference between
the first and second selectivity estimates plus one of the first or the
second selectivity estimates.

23. (Original) The method of claim 21, wherein the plurality of columns are substantially independent on each other.

24. (Original) The method of claim 23, wherein the third selectivity estimate is substantially equal to the first selectivity estimate.

25. (Original) The method of claim 21, wherein the plurality of columns are dependent on each other.

26. (Original) The method of claim 25, wherein the third selectivity estimate is substantially equal to the second selectivity estimate.

27. (Original) The method of claim 21, wherein the third selectivity estimate is within a range between the first and second selectivity estimates.

28. (Original) The method of claim 27, further comprising determining an estimate of a dependency of the columns.

29. (Original) The method of claim 28, wherein the estimate of the dependency of the columns is used to determine the third selectivity estimate.

30. (Original) The method of claim 21, wherein the third selectivity estimate is chosen to be in a central range between the first and second selectivity estimates.

31. (Currently amended) A method for optimizing a database management system process of a query, the method comprising:

collecting a plurality of single column statistics for a plurality of columns,
the plurality of single column statistics providing estimates for row
counts and unique entry counts for a single column operator;
determining a first selectivity estimate based on an assumption that the
columns are substantially independent of each other;

determining a first factor as a measure of a skew of the plurality of columns and as a measure of a dependence of a the plurality of columns;

determining a second selectivity estimate for predicates in the query using the first selectivity estimate and the first factor, the second selectivity estimate being used in optimizing processing of the query by the database management system.

32. (Currently amended) ~~The method of claim 31,~~ A method for optimizing a database management system process of a query, the method comprising:

collecting a plurality of single column statistics for a plurality of columns,

the plurality of single column statistics providing estimates for row counts and unique entry counts for a single column operator;

determining a first selectivity estimate based on an assumption that the columns are substantially independent of each other;

determining a first factor as a measure of a skew of the plurality of columns and as a measure of a dependence of a plurality of the columns;

determining a second selectivity estimate for predicates in the query using the first selectivity estimate and the first factor, the second selectivity estimate being used in optimizing processing of the query by the database management system,

wherein the first factor is determined by

computing a product of unique entry count selectivities from a sum of maximum unique entry counts for the plurality of columns,

computing a product of maximum initial unique entry counts for the plurality of columns,

computing a ratio of the product of unique entry count selectivities and the product of maximum initial entry counts,

selecting a maximum multicolumn unique entry count from multicolumn entry counts for the plurality of columns, and computing the first factor from a product of the ratio and an inverse of the maximum multicolumn unique entry count.

33. (Original) The method of claim 31, wherein the plurality of columns are substantially independent on each other.

34. (Original) The method of claim 33, wherein the second selectivity estimate is substantially equal to the first selectivity estimate.

35. (Original) The method of claim 31, wherein the plurality of columns are dependent on each other.

36. (Original) The method of claim 31, wherein the second selectivity estimate is a product of the first factor and the first selectivity estimate.

37. (New) A data processing system, comprising:
a processor;
a memory coupled to the processor;
wherein the memory stores a compiler that, when executed by the processor, determines a join selectivity value of two columns based on a first selectivity value that assumes the two columns are dependent and a second selectivity value that assumes the two columns are independent, and
wherein the compiler performs a join operation based on the join selectivity value.

38. (New) The data processing system of claim 37 wherein the compiler determines the join selectivity of two columns further based on a cross product of row counts estimated for each of the two columns.

39. (New) The data processing system of claim 38 wherein the row counts are estimated by a quantity of unique entry counts for each of the two columns.

40. (New) The data processing system of claim 37 wherein the compiler determines an intermediate selectivity value approximately halfway between the first selectivity value and the second selectivity value when a dependence between the two columns is unknown and wherein the compiler performs a join operation based on the intermediate selectivity value.

41. (New) A storage medium containing computer-readable instructions that are executable by a computer and cause the computer to:

produce a query tree based on query posed by a computer language statement;

transform the query tree into a form that represents a number of logically equivalent methods of processing the computer language statement;

estimate a cost associated with carrying out each of the logically equivalent methods,

wherein said estimate a cost comprises determining a join selectivity for two columns based on a first selectivity value that assumes the two columns are dependent and a cross product of row counts for each of the two columns.

42. (New) The storage medium of claim 41 wherein said determining a join selectivity for two columns is further based on a second selectivity value that assumes the two columns are independent.

43. (New) The storage medium of claim 41 wherein said determining a join selectivity for two columns is further based on a skew calculation that provides a correction if the two columns have different row count to unique entry count ratios.

Appl. No. 10/027,309
Amdt. dated July 29, 2004
Reply to Office action of May 5, 2004

Amendments to the Drawings:

The attached sheet of drawings includes changes to Figure 1. This sheet, which includes Figure 1, replaces the original sheet including Figure 1. In Figure 1, previously omitted element 114 has been added.

Attachment: Replacement Sheet
Annotated Sheet Showing Changes